nature research

Corresponding author(s): Douglas Black and Kathrin Plath

Last updated by author(s): May 20, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No computer codes were used to collect the data used in this study. Codes used for analyses are listed in the methods and where necessary (if not publicly available) have been posted on GitHub.
Data analysis	Statistical tests (Kolmogrov-Smirnov tests, Pearsons r, Two-tailed students t-test) were performed using standard packages available in R (used r-studio for Mac). For CLIP, RAP and ChIP data, alignments were done using the publicly available BowTie package. Pre-alignment processing for CLIP data used publicly available packages: fastq-tools, fastx-toolkit, Samtools, Bedtools, DeepTools and UCSC scripts. Peak Calls for ChIP data was done using the publicly available package MACS2. Image analysis was performed on FIJI/ImageJ. This information has been included in the methods section of the paper.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data generated during has been deposited in a public repository GEO. The account is currently private but the accession number is provided in manuscript and the

account will be made public data upon final acceptance)

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For image analyses, samples were analyzed in duplicate or triplicate (from independent experiments). Usually 50-100 units (Cells, Xist clouds) were counted per sample. This number was chosen based upon previous standards in the field and experimental measurements of the number of units needed to statistically test the observation under study.
Data exclusions	No data were excluded from this study.
Replication	1) For data assessed by microscope, experiments were repeated by the same (Amy Pandya-Jones), and different (Yolanda Markaki, Tsotne Chitiashvilli) researchers, with the same results. 2) For CLIP data, the experiments were independently repeated or matched previously published datasets from independent groups. 3) For all other experiments, reproducibility was established by having biological replicates or technical replicates (where biological replicates were not available).
Randomization	Not relevant to this study.
Blinding	For analysis of CLIP data, the data was blinded. Data was not blinded for all other experiments.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	The SI section of the submitted paper contains a table listing all this information
Validation	Most antibodies were validated using western blot. For those used in identification of the Inactive X-Chromosome, validation was done by Immunofluorescent analysis - with enrichment on the inactive X-Chromosome as a readout.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The parent F1 2-1 MS2 female ESC cell line was initially obtained from the lab of Rudoph Jaenisch where it was made, as was the v6.5 male parents ESC line. All genetic manipulations were performed by the authors as described in the methods section of the manuscript. These cell lines are not on the ICLAC cross-contamination list.
Authentication	Authentication of cell lines is not necessary as these cell lines are not on the ICLAC cross-contamination list. Genetic manipulations made for this study were confirmed by Southern blot and DNA sequencing as described in the methods section of the manuscript. In addition, cell lines in which synthetic or fusion proteins were expressed, authentication was

performed by immunoblot.

Mycoplasma contamination	Parent cell lines were tested for Mycoplasma and declared negative. The derived cell lines were not tested for Mycoplasma.
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	Sequencing Data has been submitted to GEO under record # GSE137305. While it remains in private status the following token can be used for reviewer access: izozcqsinlgldst
Files in database submission	PTBP1_iclip_Xist.bw bigWig MATR3_iclip_Xist.bw bigWig CELF1_eclip_Xist.bw bigWig CELF1_input_Xist.bw bigWig PTBP1chip_NoDox_Xist.bw bigWig PTBP1chip_Dox_Xist.bw bigWig PTBP1_iclip_Xist.fastq.gz Fastq MATR3_iclip_Xist.fastq.gz Fastq CELF1_eclip_Xist.fastq.gz Fastq CELF1_input_Xist.fastq.gz Fastq PTBP1chip_NoDox_Xist.fastq.gz Fastq PTBP1chip_Dox_Xist.fastq.gz Fastq RAP +E: 162_v_SRR850637^smoothed.bw bigWig RAP -E: 5A1_v_SRR850637^smoothed.bw bigWig RAP +E: 162_R1.fastq.gz 162_R2.fastq.gz Fastq RAP -E: 5A1_R1.fastq.gz 5A1_R2.fastq.gz Fastq
Genome browser session (e.g. UCSC)	https://genome.ucsc.edu/s/PTBP1_MATR3_CELF1/PTBP1_MATR3_CELF1

Methodology

Replicates	ChIP data was reproduced twice with close agreement - both showing a peak across the E-repeat in Xist. PTBP1 and MATR3 CLIP experiments were performed once. PTBP1 CLIP results match published reports using the similar conditions and cell lines. CELF1 CLIP was performed twice, second replicate was in done under slightly different Xist induction conditions so not included in this study, but both replicates matched closely with strong a strong CELF peak across the 5' region of the Xist E repeat. RAP-seq experiments were performed once.
Sequencing depth	PTBP1 CLIP: 3.5M unique reads, 100bp single end MATRIN3 CLIP: 5.3M unique reads, 100bp single end CELF1 CLIP: 8.9M unique reads, 51bp single end CELF1 Input 5.7M unique reads, 51bp single end PTBP1 ChIP No Dox: 11.1M unique reads, 51bp single end PTBP1 ChIP Dox: 10.5M unique reads, 51bp single end RAP +E: ~20M reads, 41 and 33bp, paired end. RAP - E: ~20M reads, 41 and 33bp, paired end.
Antibodies	For PTBP1 CLIP and ChIP experiments, abcam ab5642 was used. For the MATRIN3 CLIP experiment Abcam ab151714, clone EPR10635(B) was used. For the CELF1 CLIP experiment Abcam ab129115, clone EPR8298(B) was used.
Peak calling parameters	ChIP peaks in Xist were called with MACS2 as described in the methods section of the manuscript. CLiP peaks were not called. RAP Peaks were not called.
Data quality	Data quality was controlled for using aligner specific parameters, including excluding reads shorter than 20 nt.
Software	Basecalls performed using CASAVA version 1.4, ChIP-seq, RAP-seq and CLIP-seq reads were aligned to the mm9 or mm10 genome assembly using Bowtie2 (Langmead et al. 2012) with only those reads that aligned to a unique position with no more than two sequence mismatches were retained for further analysis. CHIP-seq Peaks were called with MACS2. CHIP-, RAP- and CLIP-seq analysis is detailed in the methods.